High-Performance Computing (HPC) and Big Data Solutions for Mobility Design and Planning

Project ID # eems037

Jane Macfarlane

June 20,2018

OVERVIEW

Timeline

Project Start: October 2017
Project End: September 2020
15% complete

Barriers

- Complexity of urban scale networks are too large to model in reasonable compute time
- Ingestion and understanding of real-world data in near real-time
- Optimization of energy productivity and mobility across complex networks

Budget

Total project funding:

$6M / 3 years

$2M per year / 4 National Labs

Partners

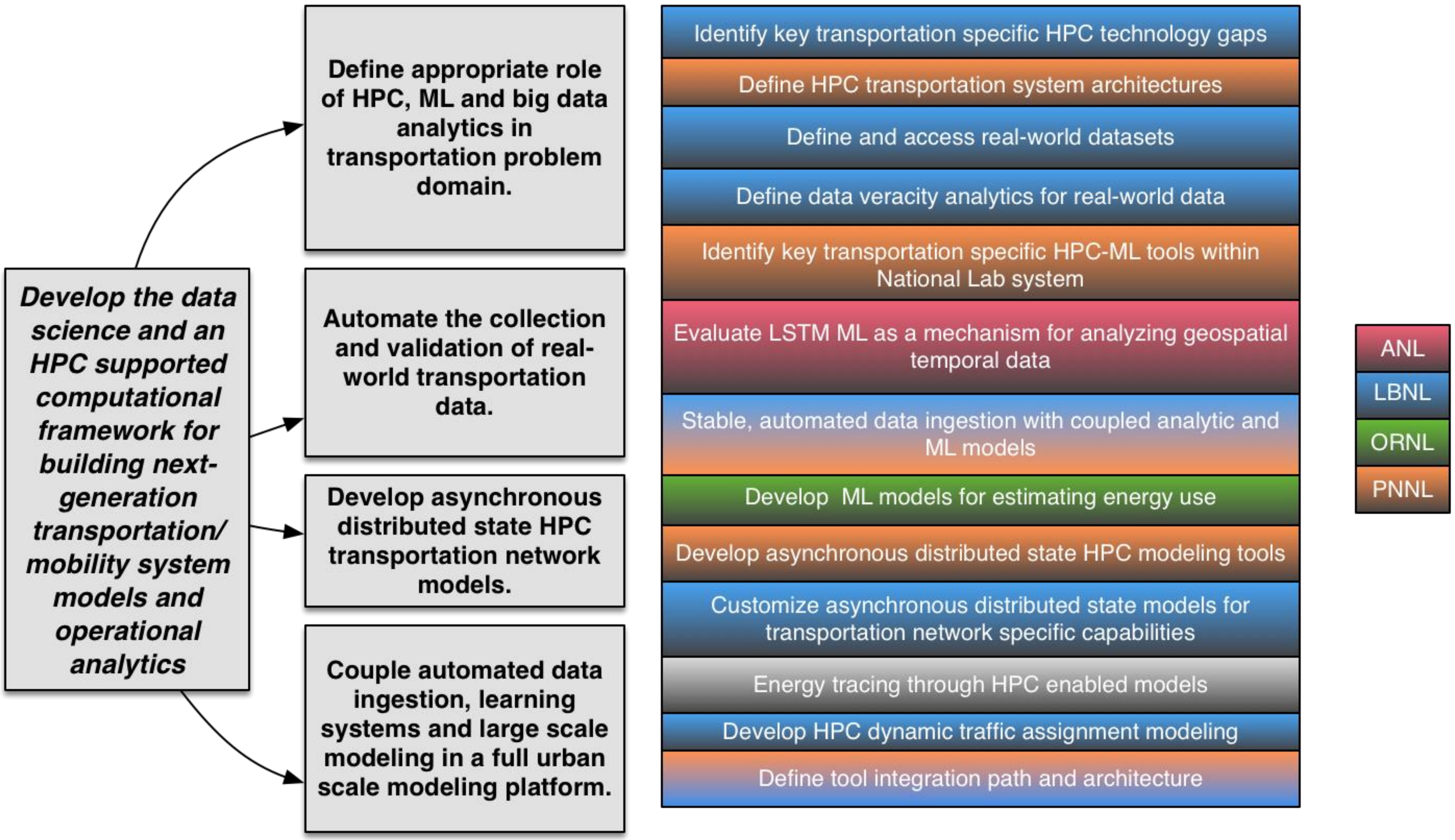
PNNL, ANL, ORNL

Connected Corridor, UCB, CalTrans

RELEVANCE

- Overall Objective: Support the EEMS program mission to create new knowledge, tools, insights, and technology solutions that increase mobility energy productivity for individuals and businesses by:
  - Developing the High Performance Computing tools to rapidly model large scale transportation networks using real-world, near real-time data.
  - Integrating energy, productivity and mobility measures to determine optimization opportunities.
- Objectives this period:
  - Identify appropriate modeling tools (e.g. deep learning) and platforms.
  - Demonstrate HPC transportation model for capturing urban scale traffic dynamics.
  - Estimate the energy cost and productivity loss of congestion.
  - Analyze real-world sensor data to estimate network demand.
- Impact:
  - Develop new control ideas for optimizing energy, productivity and mobility for normal traffic conditions and network stress conditions

WORK BREAKDOWN



MILESTONES

				FY18				FY19				Status
WBS		Task	Primary RI	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
0		**Project Coordination**	**Berkeley**									
1		**Define appropriate Role of HPC, ML, and big data analytics for transportation**	**Berkeley**									
	M1.3.1	Complete deployment of initial HPC and HPC-ML toolset as appropriate	PNNL				X					Complete
2		**Automated Transportation Data: Veracity, Velocity, and Geospatial-Temporal**										
	M2.1.2	Establish organizationally efficient data access process	Berkeley					X				On Track
	M2.2.2	Select initial HPC-ML tools	PNNL			X						Complete
	M2.3.1	Select initial training dataset from Connected Corridor	ANL			X						Complete
	M2.3.2	Demonstrate viability of LSTM neural architecture for G-T data	ANL					Go/NG				On Track
	M2.5.2	Define path for integration of energy models into HPC framework	ORNL				X					Complete
	M2.5.3	Evaluate efficacy of ML approach and impact of data veracity on energy estimates	ORNL						Go/NG			On Track
3		**Develop asynchronous distributed state HPC transportation network models**	**Berkeley**									
	M3.2.2	Demonstrate asynchronous distributed state model on large scale network	Berkeley				Go/NG					Satisfied
	M3.3.2	Develop data transfer approach to integrate optimization models							X			On Track
	M3.3.3	Demonstrate HPC dynamic traffic assignment (DTA) modeling									X	On Track
4		**Couple automated data ingestion, learning systems and large scale modeling**										
	M4.1.1	Preliminary dataflow diagram						X				On Track
	4.2.1	MaTEx deployed at all labs	All				X					Complete

APPROACH : Mobility Modeling & Optimization

Connected Corridor, Los Angeles
UCBerkeley
CalTrans
LA Metro

Automated Data Ingestion
Veracity
Fusion
Model Patterns
Predictive Analytics

Input to Models
Feedback to improve ingestion

Rapid Network Simulation
Network Dynamics
Analytics
Congestion
Energy, Mobility, Productivity

Data Science
Statistics
Machine Learning
Predictive Analytics

Computational Tools

Network Modeling and Simulation
Parallel Discrete Event Simulation
Traffic Assignment

High Performance Computing

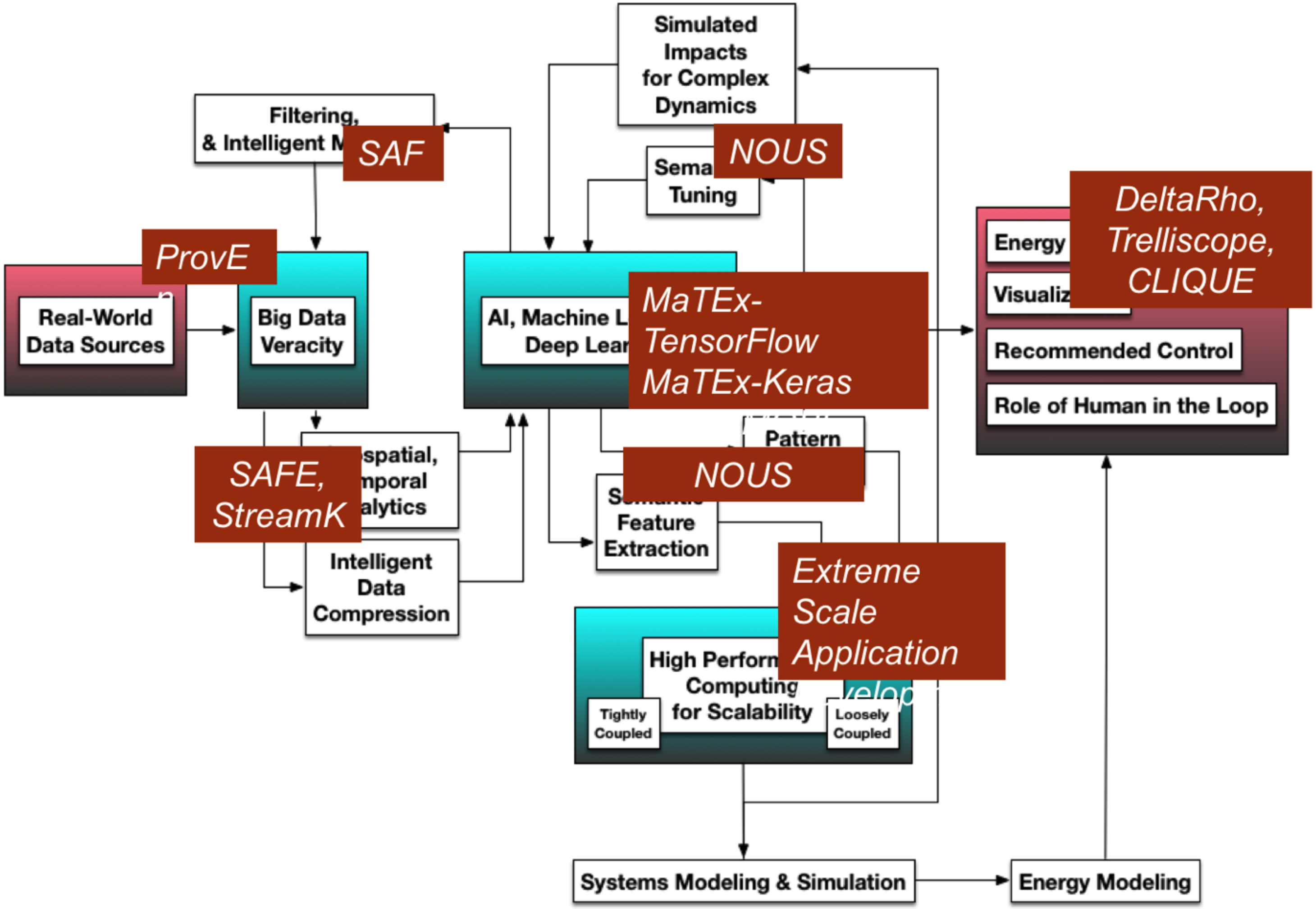
System Learnings

Real World Control and Optimization

Develop high-speed HPC enabled tools that will create actionable control predictions at the network level

Berkeley Lab, Berkeley University of California, Pacific Northwest National Laboratory, Oak Ridge National Laboratory, Argonne National Laboratory

APPROACH : Tool Layer



ACCOMPLISHMENTS:
Promising Spatial & Temporal Techniques

A multi-layered approach incorporating both temporal and spatial learning

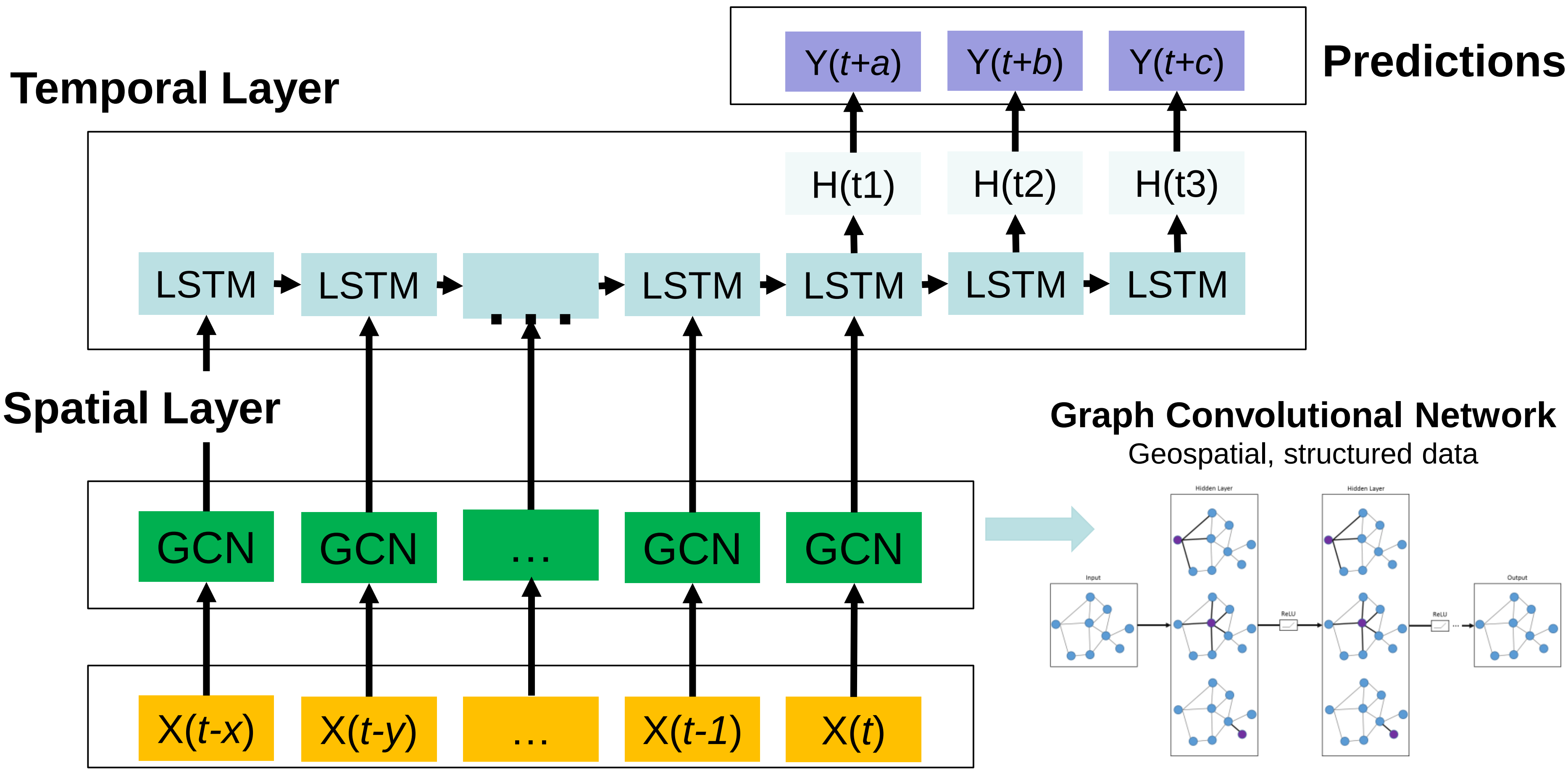
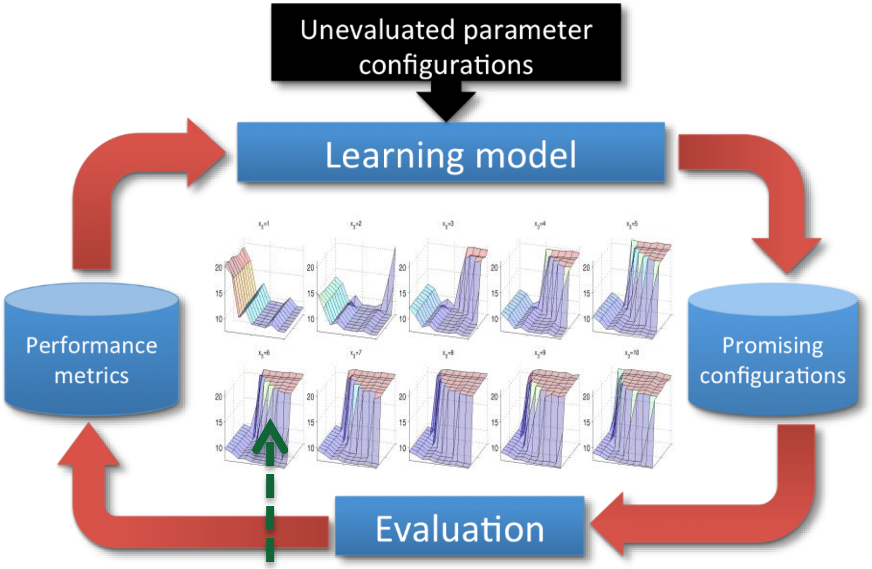


Figure adapted from: Graph Convolutional Networks, https://tkipf.github.io/graph-convolutional-networks

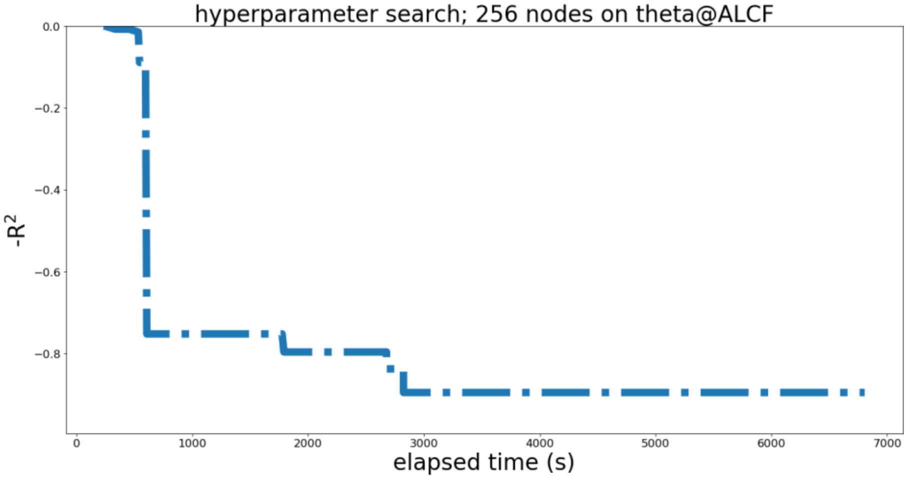
ACCOMPLISHMENTS:
Hyper-parameter Search for LSTM

LSTM Traffic POC on Theta at Argonne's ALCF

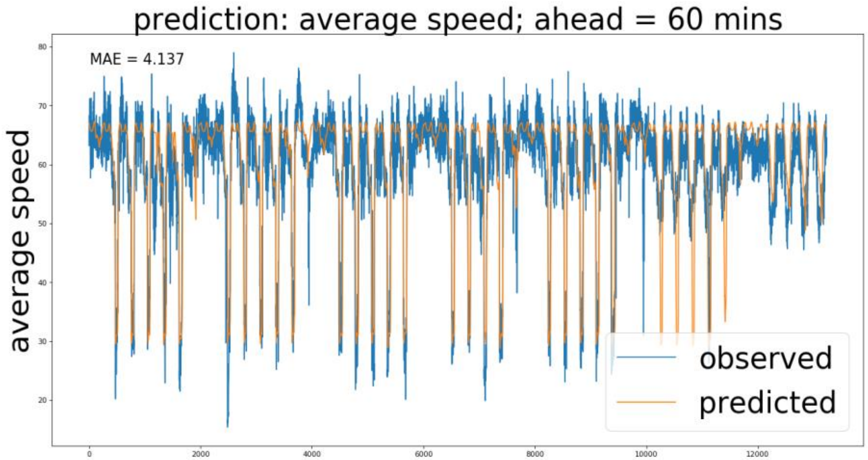
- LSTM hyper-parameter tuning POC running at increased scale
- 256 nodes; each node evaluates a deep neural network

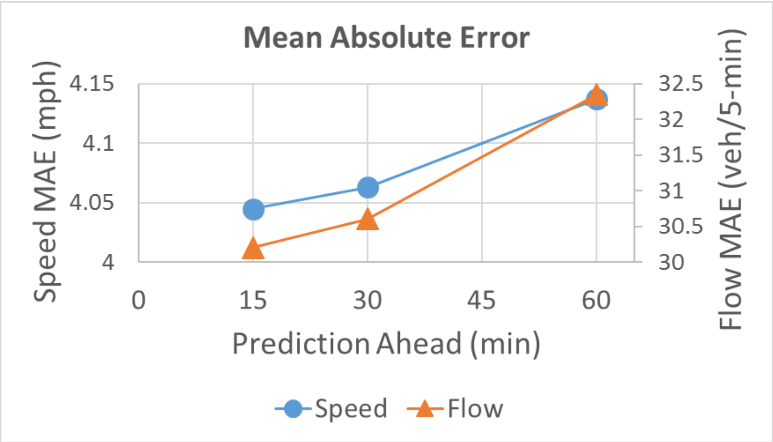


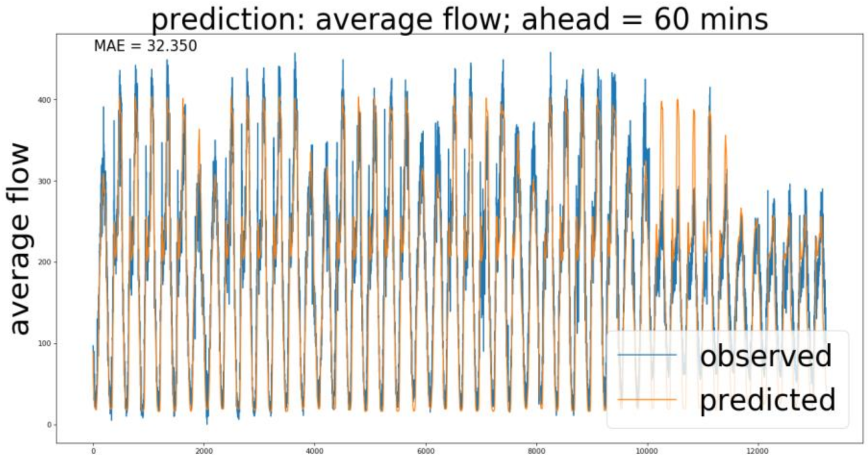
Example Surrogate Model Fitted to Sampled Performance
(iterative refinement improves the learning model)

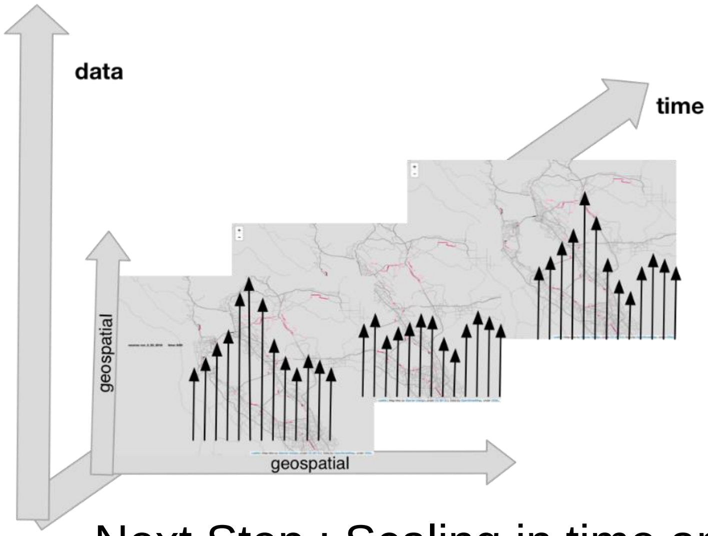


ACCOMPLISHMENTS: Single Station LSTM Based Speed & Flow Estimation POC









Next Step : Scaling in time and space

ACCOMPLISHMENTS:
Actor-Based Traffic Model Representation

- Actor-based model: links are actors, vehicles are propagated through the system via events passed between actors
- Event signals the arrival of a vehicle at a link at some time T_0
- Link actor mediates the congestion experienced by each vehicle traversing the link
- Link computes vehicle traversal time and schedules an event for the vehicle's arrival at the next link actor in the vehicle's path at time T_1

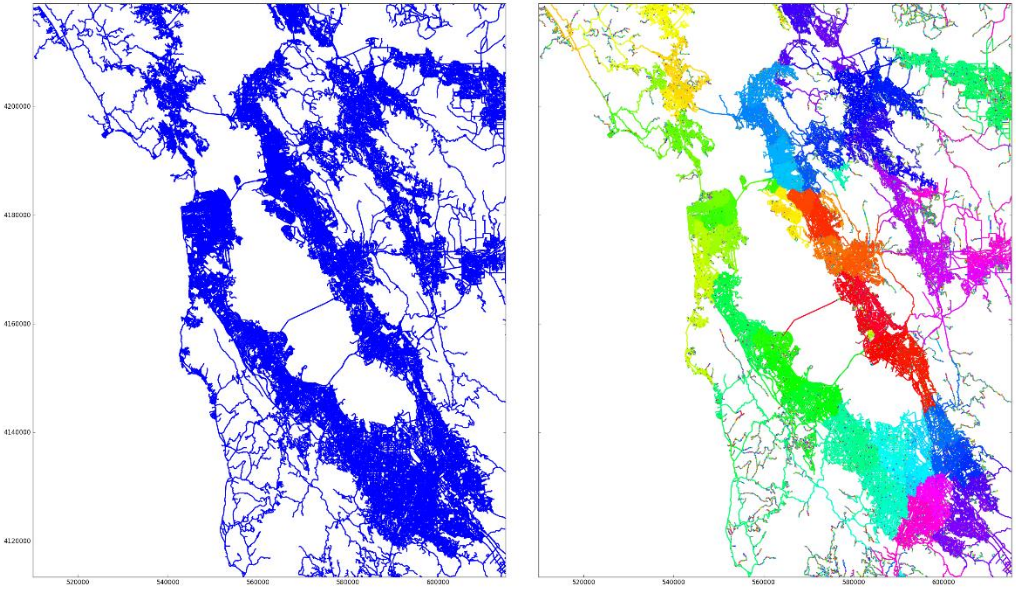
$$S_a\left(v_a\right) = t_a\left(1 + 0.15\left(\frac{v_a}{c_a}\right)^4\right)$$

Send Event: $T_1 = T_0 + S_a(v_a)$ **Link Actor** Receive Event: T_0

freehandz / 123RF Stock Photo

ACCOMPLISHMENTS:
Optimistic Parallel Discrete Event Simulation

- Simulation is parallelized by splitting links across multiple computer nodes/processes/threads to logical processors (LPs)
- Vehicles traverse between LPs and must be communicated across the network



Intelligent Geospatial Partitioning of the Network Graph

Conservative PDES:
Requires every rank to be synchronized to a global time step

Optimistic PDES:
Allows ranks to execute without synchronization and enforces causality by rolling back mis-speculatively executed events
Reduce communications and rollback by multi-objective partitioning of link actors based on event load

ACCOMPLISHMENTS: Urban Scale Flow Modeling

APPROACH: Four Part Vehicle Energy Model

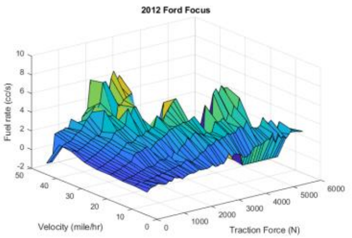
Objective: Estimate/predict fuel (energy) consumption for each vehicle on each path segment.

- Determine resistant force $F_r(v)$,
- Determine the vehicle traction force, $F_t(v)$
- Determine fuel rate characteristic map, $c(v) = f(F_t(v), v)$, using ANL datasets
- Compute fuel consumption rate for each vehicle based on the velocity in Mobiliti

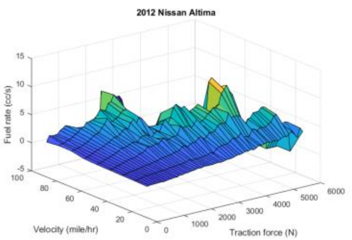
Dynamometer Test Data (ANL) → Data fitting and tuning → 2D Lookup Table

Timestamp [sec]	Dyno Speed [mph]	Dyno Tractive Effort [N]	Fuel flow [cc/s]
38.1	10.486429	112.832683	0.37709
38.2	10.488464	132.010357	0.392381
38.3	10.487446	135.984109	0.420566
…	…	…	…

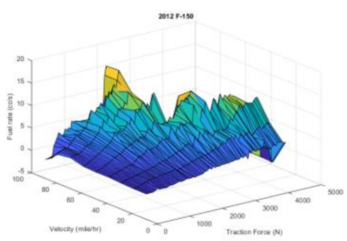
$$c = f(F_t(v), v)$$



2012 Ford Focus



2012 Nissan Altima



2012 Ford F-150

ACCOMPLISHMENT: Four Part Vehicle Energy Model

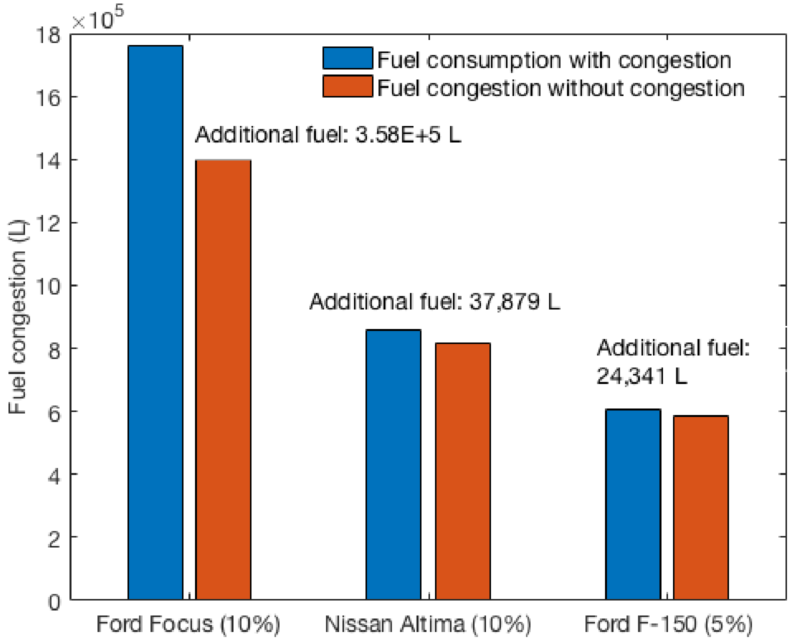
Compute fuel consumption rate for each agent:

For the trip n, the total fuel consumption $e_n = \sum_{i \in Path} f\left(F_t\left(v_n^i\right), v_n^i\right) \cdot \frac{d_i}{v_n^i}$

For the road link i, the accumulated fuel consumption $E_i = \sum_{n \in N_i} f\left(F_t\left(v_n^i\right), v_n^i\right) \cdot \frac{d_i}{v_n^i}$

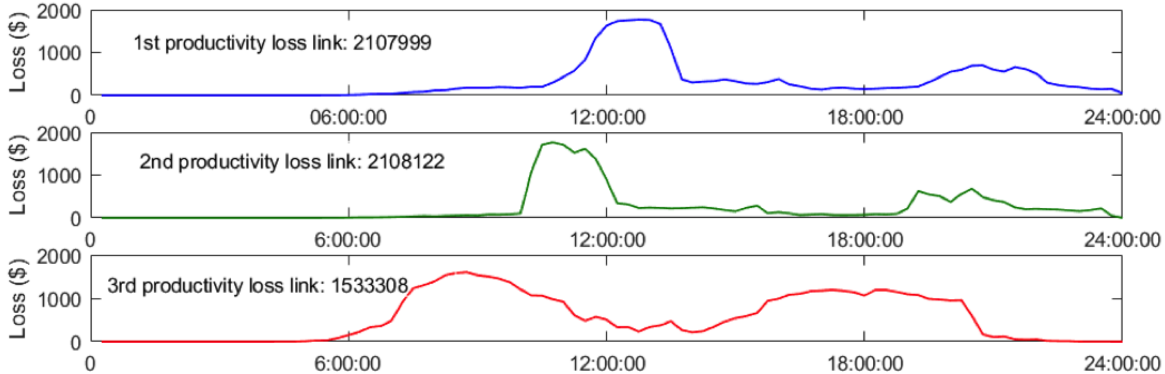
where d_i is the length of road link i, v_n^i is the velocity of trip n through link i, N_i denotes all the trips through link i.

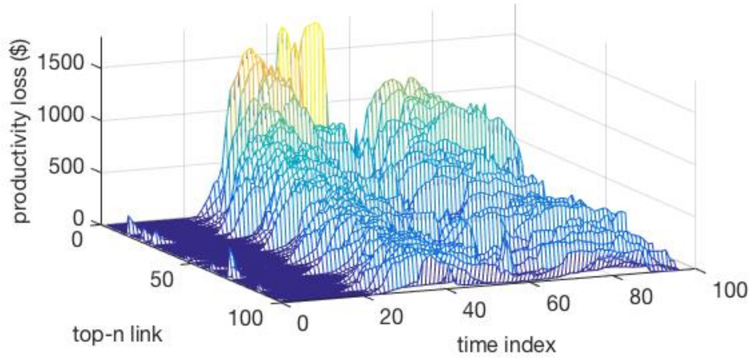
Vehicle Model	% of Fleet	Replication Factor
2012 Focus	10%	
2012 Altima	10%	8
2012 F-150	5%	

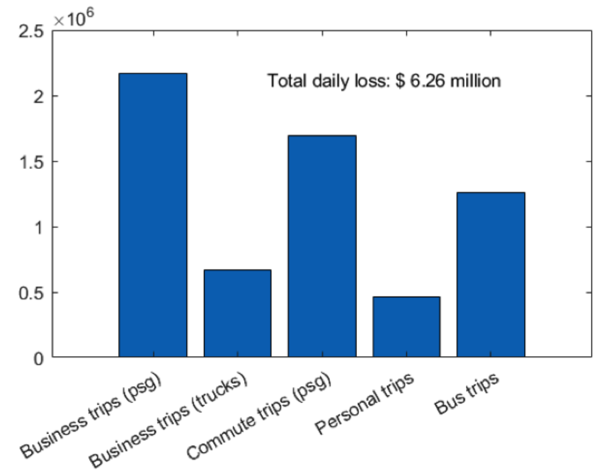


BERKELEY LAB | Berkeley UNIVERSITY OF CALIFORNIA | Pacific Northwest NATIONAL LABORATORY | OAK RIDGE National Laboratory | Argonne NATIONAL LABORATORY

ACCOMPLISHMENTS: Urban Scale Productivity Impact







- Time-varying loss of productivity across top links
- Up to $2000 loss per 15 mins on the top congested links
- Total daily loss is more than $6 million

BERKELEY LAB Berkeley UNIVERSITY OF CALIFORNIA Pacific Northwest NATIONAL LABORATORY OAK RIDGE National Laboratory Argonne NATIONAL LABORATORY

ACCOMPLISHMENTS: Deployed HPC Enabled Machine Learning Platform (MaTEx)

- MaTEx is a collection of parallel machine learning and data mining (MLDM) algorithms for desktops, supercomputers and cloud computing systems.

- MaTEx provides high performance implementations of DL algorithms
  - Google TensorFlow as the baseline
  - MPI for inter-node communication
  - multi-threading/CUDA (cuDNN) for intra-node execution

MaTEx-TensorFlow Code

```
import tensorflow as tf
import numpy as np
...
from datasets import DataSet
...
image_net = DataSet(...)
data = image_net.training_data
labels = image_net.training_labels
...
# Setting up the network
...
# Setting up optimizer
...
init = tf.global_variables_initializer()
sess = tf.Session()
sess.run(init)
...
# Run training regime
```

Original TF Code

```
import tensorflow as tf
import numpy as np
...

...

data = ...    # Load training data
labels = ... # Load Labels
...
# Setting up the network
...
# Setting up optimizer
...
init = tf.global_variables_initializer()
sess = tf.Session()
sess.run(init)
...
# Run training regime
```

Supports automatic distribution of HDF5, CSV, PNetCDF formats

TensorFlow

K + TensorFlow

Caffe

MaTEx

NERSC

Argonne

OLCF

1) Open source software with users in academia, laboratories and industry
2) Supports graphics processing unit (GPU), central processing unit (CPU) clusters/ LCFs with high-end systems/interconnects
3) Machine Learning Toolkit for Extreme Scale -MaTEx: github.com/matex-org/ matex

- MaTEx also supports
  - K-means
  - Spectral Clustering algorithms for clustering
  - Support Vector Machines
  - KNN algorithms for classification
  - FP-Growth for Association Rule Mining.

COLLABORATION AND COORDINATION

National Laboratories : HPC Modeling

Government and Academia : Infrastructure Data

Connected Corridor, University of California Berkeley
CalTrans, LA Metro, LA DOT

Industry : Mobility Data

GPS Data for Connected Corridor Region

TAZ Movement Data / Validation

REMAINING CHALLENGES AND BARRIERS

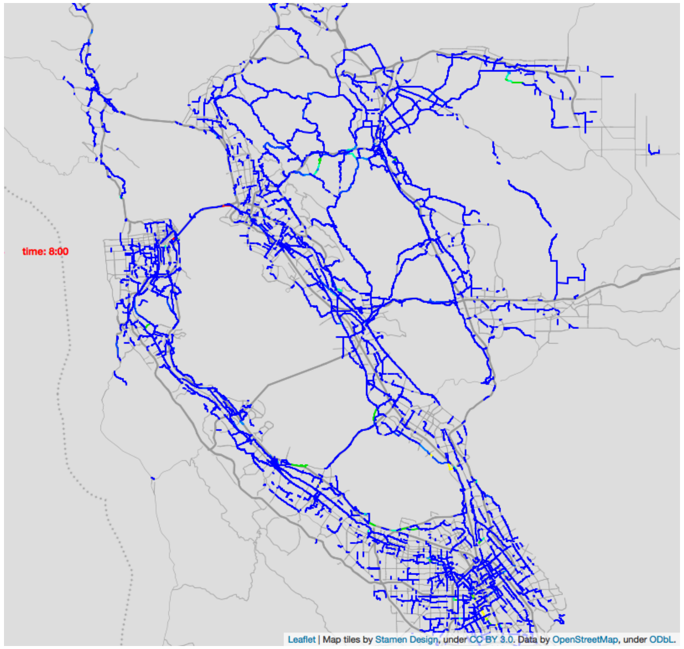
- Acquisition of urban scale mobility data
- Development of learning models for sparse and low quality data sets
- Data fusion mechanisms for creating improved demand models from real-world data
- Extension of small scale models to full urban scale models
- Development of learning models for large scale networks

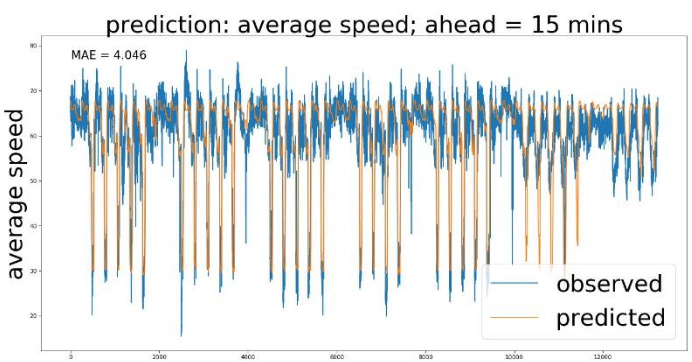
FUTURE RESEARCH

- Characterizing real-world mobility demand using automated, data-fusion ML models
- Use of simulator to create emergent behaviors:
  - Extend models and simulator to investigate impact of more intelligent transportation system agents – dynamic routing.
  - Investigate impact of induced agent rerouting as a result of a network change
- Impact of routing and control algorithms on energy, productivity and mobility measures with more sophisticated network representation, eg. improved map information
- Improved performance of the simulator to generate data sets for machine learning algorithms for the purpose of creating large scale network characterizations
- Investigate the use of resultant models in real-time decision making

Any proposed future work is subject to change based on funding levels

SUMMARY





- Urban-scale model for investigating impact of mobility dynamics running on HPC
  - Capable of running large network and demand (2M links, 7M agents) in minutes
  - Energy model under development
  - Economic cost model in place
- Machine Learning approach to estimating speed and flow from real-world data - allows for understanding and modeling true dynamics
- HPC enabled tool chain in place to create integrated problem solutions

BERKELEY LAB Berkeley UNIVERSITY OF CALIFORNIA Pacific Northwest NATIONAL LABORATORY OAK RIDGE National Laboratory Argonne NATIONAL LABORATORY

Technical Back-Up Slides

APPROACH: Functional Layer with Analytics